DOCUMENT RESUME

ED 223 514 SO 014 384


AUTHOR Colby, Anne; Kohlberg, Lawrence

TITLE Invariant Sequence and Internal Consistency in Moral Judgment Stages.

INSTITUTION Harvard Univ., Cambridge, Mass. Graduate School of Education.


PUB DATE Dec 81
NOTE 18p.; For a detailed report of this research see SO 014 382.
PUB TYPE Reports - Research/Technical (143) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Developmental Stages; Longitudinal Studies; Males; *Moral Development; Moral Values; *Research Design; *Research Methodology; *Test Reliability; *Test Validity; Value Judgment
IDENTIFIERS Kohlberg (Lawrence)



ABSTRACT

A 20-year study to monitor moral developmental stages is summarized with a focus on the invariant sequence and internal consistency of each stage. It is theorized that an individual passes through several stages in attaining moral judgment. As children grow older, they are able to integrate diverse points of view on a moral conflict. Each stage presupposes understanding gained at previous stages. This analysis of the stage employs a new scoring method, Standard Issue Scoring, based on a redefinition of the stages from a sociomoral perspective. Fifty males were interviewed in 3- to 4-year intervals; variables included age, social class, and sociometric status. Results indicated that the instrument is highly reliable. The two criteria for judging validity were invariance of stage sequence (stages were not skipped nor did individuals regress from a higher to a lower stage). Also, individuals within each stage demonstrated internal consistency; that is, their judgment was consistent throughout several moral issues or dilemmas. The data supported the validity of both of these measures. (KC)



**********
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
**********

ED223514

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☐ This document has been reproduced as received from the person or organization originating it.

☒ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
Lawrence Kohlberg
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

# Invariant Sequence and Internal Consistency in Moral Judgment Stages

by

Anne Colby and Lawrence Kohlberg

December, 1981

SP 014 384

# Invariant Sequence and Internal Consistency in Moral Judgment Stages

Anne Colby and Lawrence Kohlberg

In this paper we will present the results of a twenty year longitudinal study of moral development. The study represents an attempt to document the basic assumptions of Kohlberg's account of moral judgment development.

This approach to moral judgment focuses on the qualitative form of the child's moral reasoning and on developmental changes in that reasoning. Kohlberg has attempted to describe general organizational or structural features of moral judgment that can be shown to develop in a regular sequence of stages. The concept of structure implies that a consistent logic or form of reasoning can be abstracted from the content of an individual's responses to a variety of situations. It implies that moral development may be defined in terms of the qualitative reorganization of the individual's pattern of thought rather than the learning of new content. Each new reorganization integrates within a broader perspective the insights that were achieved at lower stages. The developing child becomes better able to understand and integrate diverse points of view on a moral conflict situation and to take more of the relevant situational factors into account. In this sense, each stage presupposes the understanding gained at previous stages. As a result, each stage provides a more adequate way of making and justifying moral judgments. The order in which the stages develop is said to be the same in each individual not because the stages are innate, but because of the underlying logic of the sequence.

Kohlberg has hypothesized that the developmental levels that he has described constitute stages in a strict Piagetian sense. First, the stage concept implies that under normal environmental conditions developmental change will always be upward in direction. Second, it implies that there will be no stage skipping. The individual must pass through each stage in order to reach the next stage in the sequence. Third, the stage concept implies that individual's thinking will be at a single dominant stage across varying content, though use of the stage adjacent to the dominant stage may also be expected. The results of previous studies of moral judgment have not completely supported the strong stage claim, as critics like Kurtines and Grief (1974) have pointed out. In part, the ambiguity in some of these findings has been due to the limited reliability of Kohlberg's 1958 method of interviewing and scoring moral stage. The longitudinal data published by Holstein (1976), Kuhn (19 ), and White et al (1978) have shown anomalies in stage sequence. An earlier analysis of the first ten years of Kohlberg's data also showed a substantial number of sequence anomalies (Kramer 1968, Kohlberg and Kramer, 1969). It has not been clear whether these anomalies represent a failure of fit of the strict stage model to moral judgment development, whether they have represented confusions in the conceptual definitions of the stages, or problems in the reliability or validity of the measure. The study described here is an attempt to address the validity of the stage model as it applies to longitudinal data and the associated problems of stage definition and measurement.

The study includes a reanalysis of Kohlberg's 1956-1968 data along with analysis of the subsequent data collected from the same subjects from 1968

through 1976. The current analysis involved the application of a new scoring method (Standard Issue Scoring) based on a substantially revised account of the stages. One especially noteworthy change in moral judgment stage scoring is the omission of Stage 6 from the current manual. Stage 6 was omitted partly because none of the interviews in the longitudinal sample seemed intuitively to be Stage 6, partly because the standard dilemmas are not ideal for differentiating between Stages 5 and 6. The question of whether Stage 6 should be included as a natural psychological stage in the moral development sequence will remain unresolved until research using more appropriate moral dilemmas and interviewing techniques is conducted with a special sample of people likely to have developed beyond Stage 5.

Before turning to the study, let us consider briefly the development of the coding scheme. In the early 1970's Kohlberg and his colleagues began to redefine the central features of the moral judgment stages and to construct a more adequate assessment instrument. Seven of the longitudinal cases were used as the data base for this redefinition. It was expected that this process would yield a more precise and accurate picture of moral judgment development than did the earlier stage descriptions. For example, it was expected that when Kohlberg's remaining longitudinal cases were analyzed using the revised stage criteria, the data would more closely fit the core assumptions of cognitive-development theory than they did in Kramer's original analysis. Since the longitudinal interviews used to validate the reformulations were kept entirely separate from those used to generate them, the revision process avoided circularity by allowing a test of prediction to data other than those used to generate the scoring system.

The early scoring systems, Sentence Rating and Global Story Rating, were based on what was essentially a content analysis. That is, both systems focused on what concerns a subject brought to bear in resolving a dilemma (for example, a concern for love between a husband and wife, the importance of obeying the law, a fear of punishment) and treated those concerns as indicators of developmental stage. These scoring systems yielded enough sequence anomalies in the 1968 analysis of Kohlberg's longitudinal data to warrant substantial revision of the stage definitions. This revision process resulted in a clearer differentiation of moral judgment structure from content. That is, more formal or abstract features of moral judgment were identified and formed the core of the new stage conceptions.

The basic developmental concept underlying the revised stage sequence is level of sociomoral perspective, the characteristic point of view from which the individual formulates moral judgments. In regard to level of sociomoral perspective, we interpret the perspective-taking underlying the moral stages as intrinsically moral in nature rather than a logical or social-cognitive structure applied to the moral domain. In this interpretation we follow Turiel (19 ) and Damon (in press) in their contention that there are many types of perspective-taking, each of which develops somewhat separately as a result of experience in a particular domain. In this view, spatial, social, and moral perspective-taking are fundamentally different processes rather than applications of a single general structure to different content areas. That is, the form of spatial perspective-taking is intrinsically spatial, not moral or social. The form of moral perspective-taking on the other hand in intrinsically moral.

According to this formulation, the most basic and general defining

feature of moral judgment development is development through the levels of sociomoral perspective. These levels provide a general organization of moral judgment and serve to inform and unite other more specific moral concepts such as the nature of the morally right or good, the nature of moral reciprocity, of moral rules, rights, obligation or duty, of fairness, welfare consequences, and moral values such as obediance to authority, preservation of human life, and maintenance of contracts and affectional relations. Within each of these specific moral concepts, the form of developmental change is to some extent specific to the nature of the particular concept in question. However, the general moral perspective can be seen to underlie its more specific manifestations. The current scoring system, Standard Issue Scoring, is based on the stage redefinitions that center on the idea of sociomoral perspective. Its aims were to achieve greater validity through the redefinition of structure and content, to achieve greater objectivity and reliability in scoring by specifying clear and concrete stage criteria, and to define the developmental sequences of the specific "moral concepts" within each stage as well as the sequence of the global or general stage structures.

The design of the original cross-sectional study which was later followed up longitudinally included three variables: age, social class, and sociometric status. Unfortunately, gender was not a variable. All the subjects were boys. These originally cross-sectional subjects, stratified by three levels of age (10, 13, and 16) and two levels of social class and sociometric status, were followed longitudinally at regular three to four year intervals for twenty years. There were 53 boys in the longitudinal study.

At each testing time subjects were interviewed on the nine hypothetical

moral dilemmas making up the three forms of Kohlberg's moral judgment interview: Forms A, B, and C. Interviews were scored according to the Standard Issue Scoring Manual (1983). Because the number of interviews to be scored was so large, a different rater scored each of the three forms A, B, and C. Scoring of all dilemmas was done blind. That is, raters coded the responses to (or scores on) other dilemmas at the same testing time or responses to (or scores on) any of the dilemmas at other testing times.

Reliability

Reliability data of several types were compiled for Standard Issue Scoring. The results indicate that the instrument is highly reliable.

Correlation coefficients for test-retest reliability are in the high 90's. Since the correlations could be very high without much absolute agreement between scores at Time 1 and Time 2, we have also calculated percent agreement figures.

On test-retest interviews conducted about a month apart almost all subjects received scores within one third stage of each other. About three quarters received identical scores on the two interviews when a nine point scale is used, and between one half and two-thirds receive identical scores with a more finely differentiated thirteen point scale. When scores do change from Time 1 to Time 2, the change is as likely to be negative as positive so cannot be attributed to a practice effect. There were no age or sex differences in test-retest stability.

In regard to inter-rater reliability, percent agreement figures were very similar to the percent agreement figures for test-retest reliability. The correlation coefficents for inter-rater reliability were in the high 90's.

Percent agreement between forms was comparable to test-retest and inter-rater reliability: 100% of the interviews were given scores with one third stage of each other for the two forms, three-quarters received identical scores for the different forms using a 9 category system, and two-thirds received identical scores for the different forms using a 13 category system. The correlations between moral maturity scores for alternate form reliability were in the mid to high 90's.

Validity

As we have argued elsewhere (Kohlberg, 1980; Colby, 1978), the appropriate validity concept for a developmental measure such as Standard Issue Scoring is construct validity, not prediction to an external criterion. For a measure of moral judgment stage, the two most critical empirical criteria of construct validity correspond to the two most central theoretical assumptions of the stage construct. They are invariance of stage sequence and "structural wholeness" or internal consistency (generality of stage usage across moral issues or dilemmas). The results to be presented here confirm both invariant sequence and internal consistency. Among other things, we interpret "construct validity" to mean the fit of the data obtained by means of the test to primary components of its theoretical definition. The primary theoretical definition of structural moral development is that of an organization passing through invariant sequential stages. In other words, positive results of the longitudinal analysis support not only the theoretical assumptions, but also the validity of the measure. Of course negative results could be due either to a faulty theory, a faulty test, or both.

Results

Invariant stage sequence. According to our theoretical assumptions, the

developmental sequence should be identical in every individual studied. Except under extreme circumstances, there should be no deviations from perfect sequentiality. Since we have not constructed an error-free measure, however, some discrepant scores must be expected to result from measurement error. A reasonable estimate of the number of such deviations attributable to measurement error can be derived from analysis of test-retest reliability data. That is, we can assume that virtually none of our subjects will have changed in stage of moral judgment within the short test-retest interval. Any differences between scores at Times 1 and 2 can be attributed to measurement error. Therefore, our analysis of longitudinal sequence involved a comparison of the frequency of sequence deviations (or downward stage movement) in the longitudinal data with the frequency of negative Time 1 - Time 2 changes in our test-retest data. (Of course, this comparison depends upon a psychometrically adequate level of test-retest reliability.) If sequence deviations exceed test-retest instability, we cannot consider our data to support the invariant sequence assumption. A comparison with downward stage change in test-retest data for Forms A and B indicates that. Since in every case the test-retest reversals are well over twice as great as the longitudinal reversals, it seems reasonable to attribute the violations of longitudinal sequence to measurement error. The frequency of sequence violations was 4% in the longitudinal data.

In addition to predicting that at no time will a subject move downward in the developmental sequence, cognitive-developmental theory holds that each stage is a prerequisite for those that follow it. This is, the concept of invariant stage sequence implies that no stage will be omitted as development

proceeds. Since within a four year interval a subject might enter and leave a stage, we could not necessarily expect our data to corroborate this aspect of sequentiality. Fortunately, however, the interval seems to have been short enough in relation to our subjects' rate of development to capture each stage in the sequence for each subject. In fact, in no case on any form of the interview did a subject reach a stage in the sequence without having gone through each preceding stage. For the most part, changes across the four year intervals were less than a full stage.

Internal consistency. According to cognitive-developmental theory, the logic of each stage forms "a structured whole". In line with this assumption, one would expect to find a high degree of internal consistency in stage scores assigned, at least within those units that are conceptually and psychologically coherent. The data support this assumption as clearly as they do that of invariant sequence.

One indication of degree of internal consistency in moral judgment is provided by distributions for each subject of proportion of reasoning scored at each of the five stages. Our analysis of these distributions showed that most interviews received all of their scores at either a single stage or at two adjacent stages. The mean percentage of reasoning at the individual's modal stage was 68% from Form A, 72% for Form B, 69% for Form C, and 67% for Forms A, B, and C combined. The mean percentage of reasoning at the subject's two most frequently used stages (always adjacent) was 98% for Form A, 97% for Form B, 99% for Form, C, and 99% for Forms A, B, and C combined. (It should be noted that there there are three dilemmas per form and each dilemma was scored without knowledge of responses to the other dilemmas, so these

figures cannot be an artifact of scorer bias.)

Factor Analyses

Theoretical expectations lead us to believe that moral judgment development is a single general domain cutting across verbal dilemmas and issues. To examine this hypothesis further a factor analysis of scores on the 18 issues across the nine dilemmas was performed. In summary, for individual age groups and for the sample as a whole, no more than one interpretable factor emerged. In all cases, the proportion of variance accounted for by the first factor far exceeded those of succeeding factors. This implies that moral judgment as measured by Standard Interview Forms A, B, and C and scored using Standard Issue Scoring is a single, general domain.

Relation of moral judgment stage to age

As one would expect of a developmental variable, our data show a clear relationship betwen age and moral judgment stage. The correlation between age and MMS was .78. With regard to age norms in our sample, the interviews of most ten year-olds were scored at Stage 1/2 or 2, a few were scored at 2/3. Most early adolescents (13-14) were 2/3, though some were Stage 2 or 3. Almost half of the late adolescents were scored as Stage 3, about one-quarter had begun the transition to Stage 4, and a little less than one-fifth were still in transition from Stage 2 to Stage 3. Youths in their early twenties were most likely to be in transition between Stages 3 and 4 or to be still solidly at Stage 3. About one-fifth of these subjects had completed the transition to Stage 4. Most subjects from the mid-twenties through the mid-thirties were scored as Stage 3/4 mixtures (or transitionals) with decreasing numbers at Stage 3 and increasing numbers at consolidated Stage 4. Subjects scored as post-conventional (4/5 or 5) represent about one-eight to one-sixth of the

sample from the mid-twenties on.

One important issue raised by the difference between this analysis and Kramer's earlier analysis is the question of adulthood stages. In light of the emergence of Stages 5 and 6 in adolescence in their analysis, Kohlberg and Kramer (1969) concluded that "there was no new way of thinking about the moral situations that was found in adulthood and not found in adolescence" (p. 105). Our interpretation of their finding is that the early scoring systems could not discriminate properly among superficially similar moral judgments at Stages 3, 4, and 5. This meant that when "true Stage 5" began to emerge in early adulthood, it was not recognized as a qualitatively new form of reasoning. We can now differentiate principled reasoning from superficially similar conventional reasoning and we no longer see Stage 5 being used in junior high or high school. This means that there is a new stage in adulthood and that many of our subjects continued to develop in their 20's and 30's, rather than reaching a ceiling in mid-adolescence. This result is intuitively appealing since thirty-year-olds often do seem to be more sophisticated in their moral judgment than do sixteen-year olds. However, it does raise the question of whether Stage 5 is a "naturally developing" stage in the strict Piagetian sense (cf Gibbs, 1978).

The Stage Model

Our data seem to provide strong support for the strict Piagetian stage model of development. With regard to the criterion of "structured wholeness" or internal consistency, we found that the great majority of our interviews were scored at only one moral judgment stage or at most two adjacent stages. The "structured wholeness" assumption was also supported by the high degree of alternate form and test-retest reliability, the high Cronbach's alpha, and

the results of factor analyses of issues scores.

As noted earlier, the factor analyses indicate that there was a single factor of general moral level across the domain of dilemmas and issues. The absence of issue or dilemma factors along with the absence of scatter across more than two adjacent stages seems to indicate that we have suceeded in defining a coherent moral domain united by a single underlying organizational structure. This position contrasts with that taken by Damon (1977) who argues that, at least in young children, the moral domain is broken down into a number of issues or dimensions each with its own unique structure. The two moral issues on which Damon has focused are authority and distributive justice. Although his research does not use Kohlberg's hypothetical dilemmas, it is clear that Kohlberg's dilemmas do involve both authority and distributive justice. If Damon's interpretation of the moral issues as independently organized were true at the age level addressed by our study, the factor analysis should have resulted in multiple factors rather than a single general factor of moral stage that cuts across the content of the dilemmas.

This high degree of structural coherence within each interview leads to a relatively saltatory picture of development if we compare interviews across time for any single individual. Developmental change in any given three year period seems to involve predominantly an increase in the next higher stage of reasoning along with a simultaneous decrease in the lower stage. This means that the subject is most often in transition between two adjacent stages, has given up using all earlier stages and uses no reasoning more than one stage above his or her modal stage. This contrasts with a model in which the distribution of a subject's reasoning extends across all five stages and in which development involves the gradual increase in usage of the higher

stages along with decrease (but not decrease to zero) in use of lower stages. Kohlberg's early results based on the Sentence Rating method may have appeared to support this latter model, but we now interpret those results as artifactual, that is, due to inappropriate definition of the scoring unit and a confusion beteeen moral judgment content and structure.

Our results also support the Piagetian stage model's assumption of invariant developmental sequence. With a few exceptions which can be attributed to scoring error, each of our subjects proceeded through the stages in the prescribed order, neither skipping stages nor regressing to an earlier level once a later stage had been attained.

This stage model of development has recently been criticized--in the cognitive domain by Flavell (1980) and, more directly, in the moral judgment domain by Rest (1979). Rest agrees with Kohlberg that qualitatively different forms of moral judgment can be identified and that development involves the increasing use of more advanced or sophisticated types and the decreasing use of less sophisticated types. He disagrees, however, with our claim that development proceeds through a stepwise sequence of internally consistent stages. He holds instead that individuals simultaneously use reasoning of many types and that an adequate description of an individual's moral judgment must include a quantitative account of the proportion of each type rather than a global stage designation for the individual. He interprets our finding of internal consistency as a methodological artifact and points to his own results with the Defining Issues Test as consistent with a more complex model of development.

One interpretation of the discrepancy between Rest's results and our own

derives from the fact that the DIT is measuring comprehension and preference of moral judgments made by others while the Standard Issue System is measuring an individual's spontaneous production of moral judgments in response to very open-ended questions. What we are suggesting is that the development of moral judgment as a whole (including comprehension and preference as well as spontaneous production) may be too broad a scope for what Rest calls the "simple stage model". It may be the case that development of moral judgment comprehension and preference does not follow a stage-like pattern but that spontaneous production does when assessment conditions pull for the subject's competence.

This would mean that the developing individual has one basic framework for resolving moral dilemmas and justifying those resolutions. This framework has some coherence for the individual and it undergoes developmental transformation, hierarchically integrates the insights from lower levels, and so on. On the other hand, other aspects of moral development such as the comprehension of judgments made by others may not be stage-like in this sense. In fact, given our assumptions, there is little reason to expect that subjects will comprehend statements at only one stage (or at two adjacent stages). Rather, they can be expected to comprehend all stages below their own as well as possibly a stage or two above their own. Thus we would not characterize the development of moral judgment comprehension as following a stage-like sequence. Preference for moral judgments made by others is somewhat more problematic. According to a simple stage model, how would one explain the endorsement of judgments that represent stages below one's own level of spontaneous production? One possible explanation derives from the assumption of

hierarchical integration. That is, while the stage as a whole may undergo transformation as the individual develops, many of the insights achieved at the lower stages remain valid from the higher stage perspective though they are now imbedded in a more complex and sophisticated position. For example, the Stage 3 judgment that "you ought to keep a promise because the person you promised trusts you" would be endorsed or agreed to by individuals who reason predominantly at Stages 4 and 5 as well as at Stage 3. The difference is that at the higher stages one's conception of trust and reasons for its importance have developed beyond what is available at Stage 3.

Another consideration with regard to Rest's preference data is methodological. When subjects endorse an item on the DIT, it is not clear how they understand the item. The Standard Issue Moral Judgment Interview involves probing for elaboration and clarification and pooling all responses to the dilemmas which refer to a single idea (as defined by norm and element). This system will naturally lead to greater internal consistency than assigning a score for each unelaborated statement the subject judges to have some validity. We do not interpret our finding as an artifact. It is true that according to Standard Issue Scoring rules, an interview cannot receive more than one stage (or stage transition) score for a single dilemma/issue/norm/element intersection. However, scores at all five stages may be assigned to a single issue, dilemma, or interview if interview/manual matches at those stages occur. In fact, such variance does not occur in our data even though the nine dilemmas were scored independently and by three different raters. We agree with Rest that moral judgment comprehension and preference are important aspects of moral development that are separable from spontaneous

moral judgment production and that a full developmental account of an individual's moral judgment requires scores on all those aspects as well as many others. However, we feel that a careful look at our methods and data should convince the reader of the validity of the Piagetian stage model for describing the development of spontaneous moral judgment at least within the limits of our instrument and sample.

In general, then, we interpret the results of this study as consistent with cognitive developmental theory, in particular as consistent with a stage model of development. The results in relation to moral judgment compare favorably with longitudinal studies of the stage model in purely cognitive development (Kohn, 1969). Our subjects did seem to use a coherent structural orientation in thinking about a variety of moral dilemmas. Their thinking developed in a regular way up the stage sequence, neither skipping stages nor reverting to use of a prior stage. Our results also serve to validate the moral judgment stages as operationally defined in Standard Issue Scoring and to indicate that the Standard Issue system is a reliable and valid measure of moral judgment.